Data management services of NorduGrid

Dr. A. Konstantinov*, Vilnius University, Lithuania and University of Oslo, Norway
Prof. P. Eerola, Dr. B. Kónya, Dr. O. Smirnova, Lund University, Sweden
Prof. T. Ekelöf, Uppsala University, Sweden
Dr. M. Ellert, CERN, Geneva, Switzerland
Prof. J. R. Hansen, Dr. J. L. Nielsen, Dr. A. Wäänänen, Niels Bohr Institute, Copenhagen, Denmark
Prof. F. Ould-Saada, University of Oslo, Norway

Abstract

In common Grid installations, the services responsible for storing big data chunks, replication of those data and indexing their availability are usually completely decoupled. And the task of synchronizing data is passed to either user-level tools or separate services (like spiders) which are subject to failure and usually cannot perform properly if one of the underlying services fails too.

The NorduGrid Smart Storage Element (SSE) was designed to try to overcome those problems by combining the most desirable features into one service. It uses HTTPS/G for secure data transfer, Web Services (WS) for control (through the same HTTPS/G channel) and can provide information to indexing services used in middlewares based on the Globus Toolkit^{TM}. At the moment, those include the Replica Catalog and the Replica Location Service. The modular internal design of the SSE and the power of C++ object oriented programming allows one to add support for other indexing services in an easy way.

There are plans to complement the SSE with a Smart Indexing Service capable of resolving inconsistencies hence creating a robust distributed data management system.

HISTORY

The NorduGrid project [1] including partners from Denmark, Norway, Sweden, Finland and Estonia was launched in spring 2001, with the aim of creating a Grid-like infrastructure (hereinafter mentioned simply as NorduGrid) in the Nordic countries. The Globus Toolkit^{TM} [2] developed by Globus Alliance was chosen as a base for our services. From the beginning it was mostly aimed at providing services for submission, control and execution of non-interactive computing jobs. Only a minimal data management functionality was planned.

The project was initiated by the High Energy Physics community (HEP). Hence the first significant application which ran on NorduGrid was software developed for the HEP experiment ATLAS [3]. ATLAS is supposed to produce a huge amount of raw and processed data. As a low budget project NorduGrid could not maintain a big centralised data storage facility, so we had to face the problems of maintaining a distributed data storage system.

Maintaining a Grid infrastructure available 24/7 for about 2 years has given us a lot of empirical experience and defined some priorities in a list of raised problems.

It appeared that for the distributed data management system, services which provide functionality of resolving final destination of data chunks (Data Indexing Services) are often more important than availability and performance of services storing actual data (Storage Elements). That is especially true in case of data distributed over Storage Elements of different kind. Multi-GB data transfers can easily fail because of failure of the Data Indexing Service. And experience shows that services fail at much higher rate than any sane expectation would suggest.

Another problem is that proprietary protocols limit the number of useful tools and often make it difficult to integrate them into middleware being developed. Often developers have to rely on a single external implementation which might be dropped at any moment in a favor of more promising idea.

SERVICES

Client driven Data Management

NorduGrid historically uses GridFTP compatible Data Storage services. Those are mostly based on the Globus Toolkit^{TM} API.

In the NorduGrid implementation of a GridFTP server, the data access control decisions are made upon the client's Grid identity - differently from traditional FTP servers. This currently includes the subject of a GSI certificate or attributes defined by a VOMS project [4]. Authorisation is controlled either through static rules in servers' configuration or dynamically using remotely modifiable GACL [5] descriptions.

In order to support distributed and replicated data, NorduGrid uses 2 kinds of Data Indexing Services developed by the Globus Alliance. Those are the Replica Catalog (RC) and the Replica Location Service (RLS) and were chosen mostly to keep compatibility with software developed and used by other projects also based on the Globus Toolkit^{TM}. They serve as simple passive frontends to databases and do not offer too much flexibility. Hence all

* aleks@fys.uio.no

the burden of information management is put on the client applications. That makes it a non-trivial task to keep information in Data and Indexing Services synchronised. Indeed the clients become quite complex and the system's recovery from a failure can also be very problematic.

Limited functionality of the Data Indexing Services also causes a necessity to have other services to compensate for missing functionality and therefore increasing the probability of failure.

Service driven Data Management

In order to solve some of the problems a "Smart" Storage Element (SSE) was designed and implemented. It is accessible through HTTPS and HTTPG interfaces which allows a firewall friendly network setup and yet a secure standard way of transferring data. For specific data manipulation operations, a WS approach was chosen.

The SSE is implemented in C++ with the gSOAP [6] development environment used for the WS interface. The Globus Toolkit^{TM} is used for transport-level security, authentication and delegation. C++ classes complement gSOAP structures to provide capability for multi-stream HTTP file transfers.

The SSE features direct interaction with the Data Indexing Services and so creates a self-sufficient network of services. That, for example, makes data upload operation atomic and reliable. The interaction interface is modular and support for a new Data Indexing Service can be easily added. However so far only RC and RLS are supported.

The SSE also provides data transfer on request. This allows third party data transfers without continuous intervention or control from the clients' side.

Access control is maintained on service level for administrators and per file to let ordinary users maintain they data. All rules are remotely modifiable.

The SSE has no notion of a so called Physical File Name. Also there is no tree-like or any other internal hierarchical representation of data. Instead unique identifiers from the Data Indexing Service are used to access files.

The interface to the SSE is simple and frees the client applications from significant part of interactions with the Data Indexing Service.

There are still many unsolved problems though. For example such a design requires more a complex and flexible system of trust delegation which is yet to be designed. Also the implemented WS interface so far provides only the most important operations. And interoperability with other solutions through compatible interfaces like Storage Resource Manager (SRM) and Reliable File Transfer (RFT) is also a desired feature.

This SSE has being tested during ATLAS Data Challenge 2 and so far caused no problems. Support for interaction with the SSE is integrated into all data management tools developed by the NorduGrid (Fig. 1).

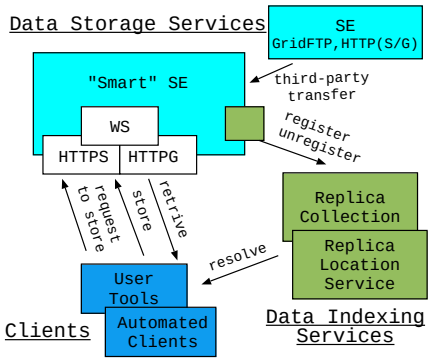

Figure 1: Interaction between the SSE and other components.

Automated Data Management

NorduGrid's plans include development of the automated data management system possibly including different solutions provided by other projects but also solutions developed internally.

Possible requirements for such a system include:

- Automatic consistency and monitoring of whole system.
- Control and initiation of data replication so providing availability of data.
- Redundancy through distribution of all parts of the system over multiple machines.
- Flexible data access control consistent through all storage endpoints.
- Make interface from client applications to the system as simple as possible.

CONCLUSIONS

NorduGrid is currently using a mixture of different services for storing and indexing data in a Grid-like infrastructure. Experience shows that such a situation does not provide a satisfactory performance. The fist steps have been made to a next generation Data Management System. As a first approach a SSE has been designed, implemented and tested. It shows good stability but does not provide all the desired functionality and has to be improved and complemented.

REFERENCES

[1] http://www.nordugrid.org/ .

[2] http://www.globus.org/ .

[3] http://atlas.web.cern.ch/Atlas/ .

[4] http://grid-auth.infn.it/ .

[5] http://www.gridsite.org/ .

[6] http://gsoap2.sourceforge.net/ .